

# The importance of reallocation for productivity growth: Evidence from European and US banking

Citation for published version (APA):

Bos, J. W. B., van Santen, P. C., & Schilp, P. (2013). *The importance of reallocation for productivity growth: Evidence from European and US banking*. Maastricht University, Graduate School of Business and Economics. GSBE Research Memoranda No. 056 <https://doi.org/10.26481/umagsb.2013056>

## Document status and date:

Published: 01/01/2013

## DOI:

[10.26481/umagsb.2013056](https://doi.org/10.26481/umagsb.2013056)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Download date: 05 May. 2023

Jaap W.B. Bos,  
Peter C. van Santen, Paul Schilp

**The Importance of  
Reallocation for Productivity  
Growth: Evidence from  
European and US Banking**

RM/13/056

**GSBE**

Maastricht University School of Business and Economics  
Graduate School of Business and Economics

P.O. Box 616  
NL- 6200 MD Maastricht  
The Netherlands

# The Importance of Reallocation for Productivity Growth: Evidence from European and US Banking

By JAAP W.B. BOS, PETER C. VAN SANTEN, PAUL SCHILP \*

*This paper quantifies the effect of reallocation dynamics on aggregate productivity developments in the banking sectors of Europe and the United States. We document an increase in productivity over the period 1995-2009, on the order of 11% in the US and 19% in Europe. At an annual frequency, under-performing banks capture market share, while more productive banks lose market share, in particular in the US. The pattern of reallocation is markedly different between the geographical regions: European productivity has grown by reallocating inputs through the first half of the sample, at the same time when reallocation diminished growth in the US. Within-firm growth has been rising steadily in both areas, largely due to technical change. The long-run positive effects of creative destruction are especially apparent in the US, where reallocation is an important driver of increases in aggregate productivity.*

JEL: O47; O30; D24; C24

Keywords: market structure, efficiency, restructuring, stochastic frontier, banking.

## I. Introduction

What drives industry growth in a consolidating market? Are all firms becoming more productive? Or do the most productive firms survive? After all, in a consolidating market, no firm's market share is guaranteed. As market concentration rises, the market restructures and scarce assets are reallocated between survivors. This implies that some firms may come out on top, and - through mergers or organic growth - capture more market share. Other firms lose market share and may eventually leave the market altogether. Yet others may see an opportunity to enter the consolidating markets. Since the seminal work by Schumpeter (1942), these effects have been well-documented

\* Bos: Maastricht University School of Business and Economics, P.O. Box 616, 6200 MD, Maastricht, The Netherlands, j.bos@maastrichtuniversity.nl. van Santen: Sveriges Riksbank, Monetary Policy Department, Research Division, SE-103 37 Stockholm, Sweden, peter.van.santen@riksbank.se. Schilp: Management Consultant, Orteliussstraat 101-2, 1057 AV Amsterdam, paul.schilp@gmail.com. We thank Jacob Bikker, Paul Cavelaars, Andrea Chegut, seminar participants at Maastricht University, Utrecht School of Economics, the Netherlands Central Bank (DNB) and participants at the 2008 North American Productivity Workshop at New York University for valuable comments. Views expressed are those of the authors and do not necessarily reflect official positions of the Riksbank. The usual disclaimer applies.

empirically.<sup>1</sup> Among the key lessons taught to us by this literature, is the notion that we need to investigate firm-level dynamics, in order to understand aggregate outcomes (Caballero et al., 1997). Theoretical models of industry dynamics emphasize selection effects at the firm level as well in shaping industry outcomes.<sup>2</sup>

A change in market structure, in particular a rapid consolidation, is often the result of a shock, such as (de)regulation, technological changes, or opening up to trade (Olley and Pakes, 1996; Melitz, 2003; Stiroh and Strahan, 2003). The timing and depth of the restructuring process depends on the timing and amplitude of the shock, and on the stage of the economic cycle (Caballero and Engel, 1993).<sup>3</sup> The reallocation itself may be smooth and rather instantaneous, but more likely it is costly and less abrupt. Exactly how costly reallocation is depends on the specificity of the industry's resources. As a result, a growing literature has studied possible reasons why resources may be more specific in some industries than in other industries (Bertola and Caballero, 1994; Mitchell and Mulherin, 1996; Caballero and Hammour, 1998). In addition, institutional, legal and other differences between countries may affect the specificity of resources, and thereby may alter reallocation dynamics (La Porta et al., 1997, 1998; Bartelsman et al., 2013). Thus, the degree of specificity may change the reallocation dynamics (Caballero and Hammour, 1996).

Reallocation dynamics, however, are not just the result of specificity, but can also be affected by regulation and other outside forces. The banking sector is amongst the most regulated and supervised industries due to its importance for economic growth and financial stability. In the US, banks have been subject to branching restrictions and interstate banking prohibitions. Similarly, banks in Europe have long been nationally oriented due to legal and institutional differences between countries. Supported by research indicating significant inefficiencies and the existence of economies of scale (Berger et al., 1995), size regulations have been relaxed and institutional barriers lowered. Both banking markets have faced significant regulatory changes, which have preceded the restructuring of these markets. As a result, the banking landscape looks rather different now compared to three decades ago: global banks supply their services around the world, while universal banks have extended the range of financial services. Naturally, expansions allow banks to exploit economies of scale and scope, which should improve welfare. At the same time, the large flows of taxpayers' money to bail out large, systemic banks during the recent financial crises has called for a re-evaluation of size restrictions. As a result, the banking sectors are expected to continue restructuring.

The literature has evaluated the consequences of banking consolidation, and

<sup>1</sup>See, amongst others, Baily et al. (1992), King and Levine (1993), Klepper (1996), Olley and Pakes (1996) and Collard-Wexler and de Loecker (2013).

<sup>2</sup>See, amongst others, Jovanovic (1982), Hopenhayn (1992), Ericson and Pakes (1995), Luttmer (2007) and Melitz and Ottaviano (2008).

<sup>3</sup>Of course, the cycle itself may change because of the shock and the resulting structural changes.

in particular mergers and acquisitions, on different measures of performance, such as cost and profit efficiency (Stiroh and Strahan, 2003), the efficient scale of production (Berger et al., 1999; Wheelock and Wilson, 1999, 2012) and productivity (Tirtiroğlu et al., 2005; Delis et al., 2011). A common theme in these papers is the study of bank-level performance measures. While useful for many purposes, in this paper, we argue that a complete evaluation should also aim to quantify the aggregate effects of restructuring. In particular, the aggregate outcomes of consolidation are different if bad-performing banks expand compared to when top-performing banks acquire more market share. To achieve this, we quantify the productivity effects of both within-firm and between-firm changes on productivity and market shares.

This approach is common in studies of the manufacturing industries where reallocation of market shares, and the entry and exit of plants, is typically found to improve aggregate outcomes in the long run (Bartelsman and Doms, 2000). Of course, the role of reallocation in driving industry productivity is not limited to manufacturing, but easily extends to banking.<sup>4</sup> We study productivity of the banking sector as it is important for the efficiency of bank lending (Wheelock and Wilson, 1999), and therefore indirectly for the financial soundness of the industry (Koutsomanoli-Filippaki et al., 2009), economic growth (Jayaratne and Strahan, 1996) and the effectiveness of monetary policy (Kashyap and Stein, 2000). We quantify the effect of reallocation dynamics on aggregate productivity developments in the banking sectors of Europe and the United States (US).<sup>5</sup>

We contribute to our understanding of aggregate productivity growth by disentangling firm-specific productivity developments from reallocation effects. We do so following three key steps. First, we model the production process of bank outputs relying on a stochastic output distance frontier model. The advantages of this approach are threefold: multiple inputs and outputs are easily handled,<sup>6</sup> output and input prices are not necessary to estimate the production function parameters, and we do not need to make assumptions about banks' objective function (Coelli and Perelman, 2000). The distance function is estimated using Call Report data for independent US banks, and BankScope data for independent European banks, covering the period 1995-2009.

Second, we compute a Malmquist productivity (TFP) index (Caves et al., 1982) for each bank-year observation. The Malmquist index is embedded in index number theory, and allows for a decomposition of productivity growth into changes in technical efficiency, returns to scale and technical progress. We use a parametric version proposed by Orea (2002), which does not require input

<sup>4</sup>See recent work by Craig and Haubrich (2013) for an example of applying labor productivity concepts to the banking sector, by studying gross loan flows.

<sup>5</sup>Throughout this paper, we use 'Europe' or 'EU' to refer to the EU-15, which includes Austria, Belgium, Denmark, Finland, France, Germany, the United Kingdom, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain and Sweden.

<sup>6</sup>This follows from the equivalence between output distance functions and the production function in the single-input case, or the transformation function in the multiple-outputs case.

and output prices, but instead computes the index as the ratio of two estimated distance functions (Shephard, 1970).

Third, using the bank-level Malmquist indices, we construct an aggregate measure of productivity in the banking sector. TFP growth is subsequently decomposed into a component measuring bank-level productivity growth, as well as a between-plant reallocation component. The latter is essentially a weighted average of bank-level changes in market shares, and therefore quantifies the effect of reallocation dynamics on aggregate productivity.

We document an increase in productivity for both European and US banking sectors over the period 1995-2009, on the order of 11% in the US and 19% in Europe. At an annual frequency, under-performing banks capture market share, while more productive banks lose market share, in particular in the US. The pattern of reallocation is markedly different between the geographical regions: European productivity has grown by reallocating inputs through the first half of the sample, at the same time when reallocation diminished growth in the US. Within-firm growth has been rising steadily in both areas, largely due to technical change. The long-run positive effects of creative destruction are especially apparent in the US, where reallocation is an important driver of increases in aggregate productivity.

We subsequently decompose within-firm growth in three components: improvements in technical efficiency, technical progress and returns to scale, all of which are identified by decomposing the Malmquist TFP index. Of these components, technical progress has been the driving force behind within-firm growth. The US banking sector is characterized by decreasing returns to scale, which restrains within-firm growth. In Europe, constant returns to scale do not have an effect on productivity.<sup>7</sup>

Finally, in an attempt to better understand reallocation dynamics and its effect on aggregate productivity, we relate reallocation to the bank size distribution and market power. In the US, large banks have, on average, lost market share to competitors, while being more productive, hence depressing aggregate growth; the reverse is true in Europe where larger, more productive banks have grown in size.

The paper proceeds as follows: Section II describes the banking sectors in Europe and the United States, with particular focus on the consolidation process. Section III presents the methodology and the data used in this paper. Section IV reports the empirical results. Section V summarizes the findings and concludes.

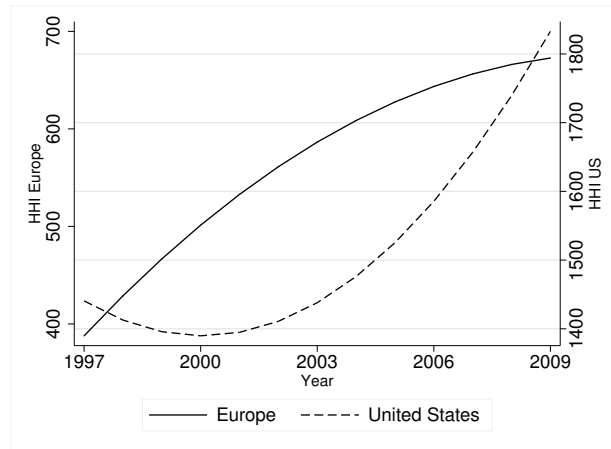
<sup>7</sup>Recent work by Wheelock and Wilson (2012) shows that US banks are characterized by economies of scale. Note that this is not incompatible with our result: only under price-taking in input and output markets do economies of scale imply increasing returns to scale. We, nor Wheelock and Wilson (2012) make this assumption.

## II. Consolidation in banking

The restructuring process of an industry is linked closely to what Schumpeter (1942) referred to as creative destruction. If the 'invisible hand' works properly, the firms with the lowest average costs survive, while the bad performers will ultimately cease to exist. On the industry level, this evolution should foster economic growth, and in the presence of scale economies will result in a more consolidated industry. Although in practice we do observe consolidation, within-sector differences can be quite large and persistent over time (Baily et al., 1992). We can distinguish between two elements that drive consolidation: market-level developments and firm-level dynamics, the latter being the subject of research in this paper.

For the former, we can point out four causes of consolidation (Berger et al., 1999). First, technological progress, e.g. internet banking, has increased the optimal scale of a bank (DeYoung et al., 2007). Second, improvements in financial conditions yielded higher profits, which freed resources available for takeovers. Third, excess capacity can be put into use for scale enlargement, reducing existing inefficiencies. Finally, to facilitate positive returns to scale, countries deregulate and loosen geographical and/or product restrictions.

Figure 1. : Consolidation in banking



*Note:* HHI is the Hirschman-Herfindahl Index, defined as the sum of the squared market shares. Market shares are calculated based on total inputs, and  $0 < HHI \leq 10000$ . Plotted are fractional polynomial fits based on calculated HHI values.

Both in Europe and in the US, banking markets have faced significant regula-

tory changes which were expected to lead to a restructuring of these markets.<sup>8</sup> For US banks, geographical restrictions have been lifted in two steps. The 1982 amendment to the Bank Holding Company (BHC) Act made it possible for out-of-state BHCs to acquire failed banks. By that time, nearly half of all 51 states allowed state-wide branching through mergers and acquisitions (cf. Stiroh and Strahan, 2003, for an overview of the years in which restrictions were lifted). The second step involved the passing of the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994, allowing nation-wide reallocation by the end of 1995 (Berger et al., 1995). The starting date of our sample (1995) allows us to consider the period immediately following these deregulations.

Similarly, Europe's national banking markets have been integrated in one market for financial services as of 1993, with the enactment of the Second Banking Coordination Directive of 1988. This Directive established the single banking license, allowing all European banks to set up branches in the entire EU area. Although Bos and Schmiedel (2007) find empirical support for a single European banking market, characterized by cost and profit meta-frontiers, Berger et al. (1999) argue that institutional and cultural differences still impede cross-border reallocation. Indeed, cross-border mergers are rare in the European Union, compared to cross-state mergers in the US. Nevertheless, the possibilities to open cross-border branches and subsidiaries exists in the European Union, and whether reallocation is hampered by the lack of mergers remains to be seen.

Both markets, after all, have experienced an increase in concentration, due to increasing merger activity, entry and exit over a period of twelve years (1997-2009). In Figure 1, we plot the development of consolidation, considering only independent banks or bank holding companies. Figure 1 shows that the Hirschmann-Herfindahl Index (HHI) of both geographical banking sectors has increased over time, indicating larger and fewer banks.<sup>9</sup> Although the overall distribution of assets is much less skewed in Europe, we observe that the overall pattern is remarkably similar, even when the US temporarily experienced a decrease in concentration after the Interstate Banking and Branching Act.

Traditionally, increases in concentration of this magnitude have been associated with decreases in competition. However, there is also evidence that increased concentration need not be related to actual competition, as documented by e.g. Bikker and Haaf (2002) and Bos et al. (2010) for the banking industry. Furthermore, on local markets (or Metropolitan Statistical Areas), which are still the relevant markets for small business and household financial services, the HHI has been constant over the last twenty years (Kwast, 2007). Park and Penacchi (2009) attribute this to the fact that banks have expanded geographically, with large banks accessing local markets by acquiring community banks. Their

<sup>8</sup>Deregulation and technological changes have been well documented for the banking industry (Rose, 1987; Berger et al., 1999; Perotti and Suarez, 2002; Wheelock and Wilson, 2000; Frame and White, 2004; Kwast, 2007).

<sup>9</sup>In constructing figure 1, we use data from the Call reports in the US, detailed in section III.C, and data from the EU reports on banking structures (European Central Bank, 2004, 2006, 2010).



model and empirical findings illustrate that small business loan rates tend to decline, indicating more competition, while deposit rates are lowered as well, mimicking less competition.

Moreover, the restructuring process is not finished. The financial crisis that started in 2007 urged politicians and policymakers to plea for a revision of the regulatory framework, which will most likely lead to a new era of tighter regulation. This will in turn lead to new restructuring incentives.

Our focus, therefore is on quantifying the productivity effects of market restructuring. What are the productivity developments in markets that experience such an increase in concentration? To what extent do all firms go through a similar process after the deregulation, and to what extent do market dynamics play a role? Do the fittest survive, or perhaps not? To answer these questions, we first need a coherent framework for studying firm and industry developments in productivity.

### III. Methodology and data

In this section, we explain how we measure the productivity of banks, and the productivity of the banking sector. We also explain how the two relate to each other, and what data we use to capture banks' production sets.

The way we typically think about productivity is as an output-input ratio. The main concept is best illustrated for a single-input, single-output firm, a feature which we generalize below. Let  $Y$  denote an output, and  $X$  denote an input, which are related using the production function  $Y_t = F_t(X_t)$ . Productivity can then be defined as  $\Pi_t = \frac{Y_t}{X_t}$ . Using this concept, productivity growth (in logarithms) can be written as:

$$(1) \quad \Delta \ln \Pi_t = \ln \frac{Y_t}{Y_{t-1}} - \ln \frac{X_t}{X_{t-1}}.$$

Suppose that, due to technical change, the production function shifts out over time (holding inputs fixed), that is,  $F_t(X_t) = \alpha_t F_{t-1}(X_t)$ , where  $\alpha_t$  denotes the growth factor. Moreover, let  $\gamma_t \in (0, 1)$  denote the productive efficiency of the firm, such that  $Y_t = \gamma_t F_t(X_t)$ . Introducing these features into our measure of productivity growth, we obtain

$$(2) \quad \Delta \ln \Pi_t = \ln \alpha_t \frac{\gamma_t}{\gamma_{t-1}} \frac{F_{t-1}(X_t)}{F_{t-1}(X_{t-1})} - \ln \frac{X_t}{X_{t-1}}$$

Finally, let  $F(X)$  be homogeneous of degree  $k$ , such that  $F_t(p_t X_t) = p_t^k F_t(X_t)$ , where  $p_t > 0$  represents the input growth factor. Using this definition in

equation (2), we obtain:

$$\begin{aligned}
(3) \quad \Delta \ln \Pi_t &= \ln \alpha_t \frac{\gamma_t}{\gamma_{t-1}} \frac{p_t^k F_{t-1}(X_{t-1})}{F_{t-1}(X_{t-1})} - \ln \frac{p_t X_{t-1}}{X_{t-1}} \\
&= \ln \alpha_t + \Delta \ln \gamma_t + (k-1) \ln \frac{X_t}{X_{t-1}}.
\end{aligned}$$

Equation (3) represents a key element of our approach. Productivity growth, defined as the growth of an output-input ratio, equals the sum of technical change, changes in technical efficiency and returns to scale in production. The remainder of this section generalizes this simple concept into a multiple-output multiple-input setting, using a more flexible representation of the production technology.

#### A. Firm productivity dynamics

We start with a model of the production of the firm. The key elements of our model include the multi-input/multi-output nature of banking and possible inefficiencies in transforming inputs into outputs. We therefore first need to define the way we capture a bank's production technology. Subsequently, we explain the way in which we estimate banks' production function. And then we demonstrate the derivation of a bank-specific productivity index.

#### TECHNOLOGY

Let  $\mathbf{x} \in R_+^J$  denote the vector of  $J$  inputs, and  $\mathbf{y} \in R_+^M$  denote the vector of  $M$  outputs. The technology<sup>10</sup> is described by the set of feasible output vectors producible from  $\mathbf{x}^t$ , that is,

$$(4) \quad \mathcal{P}^t(\mathbf{x}^t) = \{\mathbf{y}^t : \mathbf{y}^t \text{ is producible from } \mathbf{x}^t\}.$$

The outer boundary of the set  $\mathcal{P}^t(\mathbf{x}^t)$  represents the technological frontier. The output distance function measures the distance between outputs and the technological frontier, and is defined as the minimum deflator of the output vector, keeping the input vector fixed:

$$(5) \quad D_o(\mathbf{x}^t, \mathbf{y}^t) = \min_{\Psi} \left\{ \Psi > 0 : \frac{\mathbf{y}^t}{\Psi} \in \mathcal{P}^t(\mathbf{x}^t) \right\}.$$

The output distance function is non-decreasing, positively linearly homogeneous and convex in  $\mathbf{y}$ , and decreasing in  $\mathbf{x}$  (Färe and Primont, 1995). The value of the output distance function,  $D_o(\mathbf{x}^t, \mathbf{y}^t)$ , places  $\mathbf{y}/D_o(\mathbf{x}^t, \mathbf{y}^t)$  on the outer boundary of  $\mathcal{P}^t(\mathbf{x}^t)$  and on the ray through  $\mathbf{y}^t$ . Moreover, this value is

<sup>10</sup>We assume that the technology satisfies the axioms listed in Färe and Primont (1995).

the inverse of the output-oriented Farrell (1957) measure of technical efficiency, defined as the maximum feasible radial expansion of the output vector, holding inputs fixed:

$$(6) \quad \max_{\lambda} \{ \lambda \geq 1 : \lambda \cdot \mathbf{y}^t \in \mathcal{P}^t(\mathbf{x}^t) \}.$$

Making use of equation (6), and allowing for a stochastic element in equation (5), we let  $v$  denote a random error term capturing statistical noise, measurement errors and unobservable factors. The stochastic technological frontier can now be written as:

$$(7) \quad D_o(\mathbf{x}^t, \lambda \mathbf{y}^t) \cdot \exp(v^t) = 1.$$

Homogeneity of the output distance function in outputs implies that, for any scalar  $\mu > 0$ ,  $D_o(\mathbf{x}^t, \mu \lambda \mathbf{y}^t) = \mu D_o(\mathbf{x}^t, \lambda \mathbf{y}^t)$ . We impose homogeneity by setting  $\mu = \frac{1}{\lambda \cdot y_1}$ , and normalize all outputs by the first output. Hence, letting  $\tilde{\mathbf{y}} = \left( \frac{y_2}{y_1}, \dots, \frac{y_M}{y_1} \right)$  denote the  $M - 1$  vector of output ratio's, without loss of generality we can write our regression model as:

$$(8) \quad \begin{aligned} 1 &= \lambda \cdot y_1^t \cdot D_o(\mathbf{x}^t, \tilde{\mathbf{y}}) \cdot \exp(v^t); \\ -\ln y_1^t &= \ln D_o(\mathbf{x}^t, \tilde{\mathbf{y}}) + u^t + v^t, \end{aligned}$$

where  $u^t = \ln \lambda \geq 0$  denotes technical inefficiency. As in Cuesta and Orea (2002), we use the value of the distance function as our measure of technical efficiency, which is restricted to lie between 0 and 1, with greater values denoting higher efficiency. The result is a model that allows us to estimate scale economies, technical change and efficiency *while* accounting for a multi input/multi output setting. The next step involves specifying a functional form for  $D_o$ , so we can estimate the output distance frontier model.

#### SPECIFICATION

In order to estimate the model described by equation (8), we again follow Cuesta and Orea (2002) and adopt the standard flexible translog functional form for the output distance function, including time to account for non-neutral technical change. Letting  $i$  index firms and  $t$  index time, equation 8 can be

written as:

$$\begin{aligned}
(9) \quad -\ln y_{1it} = & \alpha_i + \sum_{m=2}^M \alpha_m \ln \tilde{y}_{mit} + \frac{1}{2} \sum_{m=2}^M \sum_{n=2}^M \alpha_{mn} \ln \tilde{y}_{mit} \ln \tilde{y}_{nit} + \sum_{j=1}^J \beta_j \ln x_{jit} \\
& + \frac{1}{2} \sum_{j=1}^J \sum_{k=1}^J \beta_{jk} \ln x_{jit} \ln x_{kit} + \sum_{m=2}^M \sum_{j=1}^J \delta_{mj} \ln \tilde{y}_{mit} \ln x_{jit} \\
& + \tau_0 t + \frac{1}{2} \tau_1 t^2 + \sum_{m=2}^M \tau_m t \ln \tilde{y}_{mit} + \sum_{j=1}^J \zeta_j t \ln x_{jit} + u_{it} + v_{it},
\end{aligned}$$

where symmetry requires  $a_{mn} = a_{nm}$  and  $b_{jk} = b_{kj}$ . In this specification,  $\alpha_i$  denotes a bank fixed-effect and  $u_{it}$  denotes the technical inefficiency of firm  $i$  in year  $t$ , and is assumed to be a draw from a half-normal distribution with variance  $\sigma_u^2$ . The specification in equation (9) allows for an unrestricted path of efficiency over time.<sup>11</sup> Finally,  $v_{it}$  is a normally distributed *iid* error term, independent of  $u_{it}$ , with mean zero and variance  $\sigma_v^2$ . We estimate equation (9) by maximum likelihood.

#### PRODUCTIVITY INDEX

Estimating the process with which banks transform inputs into outputs is only a first step. The next, logical step is to use the resulting production function estimations to create a productivity index. In short, we wish to know *how* each bank becomes more (or less) productive. We follow Orea (2002), and define a total factor productivity index as:

$$(10) \quad \Delta TFP_{t+1} = \frac{1}{2} \sum_{m=1}^M \left( \epsilon_m^{t+1} + \epsilon_m^t \right) \ln \left( \frac{y_m^{t+1}}{y_m^t} \right) - \frac{1}{2} \sum_{j=1}^J \left( e_j^{t+1} + e_j^t \right) \ln \left( \frac{x_j^{t+1}}{x_j^t} \right),$$

where  $\epsilon_m^t = \frac{\partial \ln D_o(x^t, y^t, t)}{\partial \ln y_m}$  and  $e_j^t = \frac{\frac{\partial \ln D_o(x^t, y^t, t)}{\partial \ln x_j}}{\sum_{k=1}^J \frac{\partial \ln D_o(x^t, y^t, t)}{\partial \ln x_k}}$ .  $\Delta TFP_{t+1}$  is thus defined as

the growth rate of outputs minus the growth rate of inputs, where outputs and inputs are weighted by their respective output distance elasticities. Note the similarity of how  $\Delta TFP$  is computed compared to equation (1) in the single-output single-input case. Furthermore, using Diewert (1976)'s Quadratic Identity Lemma, the TFP index can be decomposed into changes in technical efficiency ( $\Delta TE$ ), technical change ( $TC$ ) and returns to scale ( $SC$ ), where:<sup>12</sup>

<sup>11</sup>Our specification for inefficiency differs from Cuesta and Orea (2002), who impose a common path of inefficiency for all banks. As efficiency changes are important for our application, we prefer not to impose any common time trends.

<sup>12</sup>There are many decompositions suggested in the literature, differing with respect to how technical change and returns to scale are measured. See e.g., Färe et al. (1994), Ray and Desli (1997) and Lovell (2003).

$$(11a) \quad \Delta TE = \ln D_o(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, t+1) - \ln D_o(\mathbf{x}^t, \mathbf{y}^t, t)$$

$$(11b) \quad TC = -\frac{1}{2} \left( \frac{\partial \ln D_o(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, t+1)}{\partial t} + \frac{\partial \ln D_o(\mathbf{x}^t, \mathbf{y}^t, t)}{\partial t} \right)$$

$$(11c) \quad SC = -\frac{1}{2} \sum_{k=1}^N \left( \left( \sum_{j=1}^N \frac{\partial \ln D_o(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, t+1)}{\partial \ln x_j} + 1 \right) e_k^{t+1} \right. \\ \left. + \left( \sum_{j=1}^N \frac{\partial \ln D_o(\mathbf{x}^t, \mathbf{y}^t, t)}{\partial \ln x_j} + 1 \right) e_k^t \right) \ln \left( \frac{x_n^{t+1}}{x_n^t} \right),$$

and:

$$(12) \quad \Delta TFP = \Delta TE + TC + SC.$$

Note that both  $\Delta TFP$  as well as its components in equation (12) vary between firms and over time, and can be obtained directly from the fitted output distance function. To see why the first term measures the change in efficiency, note that, from taking logs of expression (7) in two subsequent periods,  $\ln D_o(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, t+1) - \ln D_o(\mathbf{x}^t, \mathbf{y}^t, t) = -(u^{t+1} - u^t)$ , and hence the change in the value of the output distance function measures the negative of the change in inefficiency. The second term measures the negative of the change in the distance function over time, which, by definition, equals the shifting of the technological frontier. Finally, the scale term reflects the effects of movements of the bank along the distance function, changing their input levels over time. For constant returns to scale,  $\sum_{k=1}^J \partial \ln D_o(\cdot) / \partial \ln x_k = -1$ , and hence  $SC = 0$ , while for increasing returns,  $\sum_{k=1}^J \partial \ln D_o(\cdot) / \partial \ln x_k < -1$  and  $SC > 0$ . Positive changes in technical efficiency, technical progress and increasing returns to scale are thus contributing to productivity growth. Again, we note the similarity of this decomposition with that obtained in equation (3).

Finally, to construct a measure of the *level* of productivity, we use the estimated TFP growth series and compute:

$$(13a) \quad \Pi_{it} = \Pi_{it-1} + \Delta TFP_{it}, \quad t = 2, \dots, T;$$

$$(13b) \quad \Pi_{i1} = \sum_{m=1}^M \epsilon_{mi1} \ln y_{mi1} - \sum_{j=1}^J e_{ji1} \ln x_{ji1}.$$

We use expression (13b) to compute productivity in the first period, which is

Here, we follow Lovell (2003), who recommends the above decomposition based on its ability to mimic an 'ideal' decomposition of productivity growth. Moreover, in practice, the difference between the various decompositions are minor in empirical applications.

consistent with the definition of TFP.<sup>13</sup> Now that we have thus a estimated *firm*-level measure of productivity, we can turn to *industry*-level developments of productivity.

### B. Industry productivity dynamics

To study productivity dynamics in the banking sector, we opt for a decomposition model of sector-level productivity. The goal of our decomposition analysis is to explain industry-wide growth by focusing on firm-specific behavior. The decomposition method explicitly accounts for heterogeneity in firm performance and may prove particularly insightful in restructuring industries (Foster et al., 2001, 2006). Decomposition methods are mainly used in studies of the manufacturing industries (Dunne et al., 1989, Baily et al., 1992, Olley and Pakes, 1996, Collard-Wexler and de Loecker, 2013). The decomposition allows us to differentiate between within-firm growth and growth due to creative destruction, as some firms lose market share while others gain.

Following Olley and Pakes (1996), we define industry-wide productivity as the market share weighted average of productivity levels,  $\Pi_t = \sum_{i=1}^{I_t} \Pi_{it} \theta_{it}$ .<sup>14</sup> Productivity growth between two consecutive time periods equals:

$$(14) \quad \Delta \Pi_t = \sum_{i=1}^{I_t} \Pi_{it} \theta_{it} - \sum_{i=1}^{I_{t-1}} \Pi_{it-1} \theta_{it-1}.$$

The market share of firm  $i$  in year  $t$ ,  $\theta_{it}$ , is defined as the share of total industry inputs allocated to firm  $i$ .<sup>15</sup> Over time, the number of firms may change from  $I_{t-1}$  to  $I_t$ . Between any two periods, we can identify entering firms  $N$ , present only in period  $t$ , exiting firms  $X$ , present only in  $t - 1$  and surviving firms  $S$ , which are present in  $t$  and  $t - 1$ . Hence,  $I_{t-1} = S + X$  and  $I_t = S + N$ . Rearranging equation (14), and following Foster et al. (2001), we derive the following decomposition of productivity growth, the derivation of which is presented in Appendix A:

<sup>13</sup>However, when using a firm fixed effect in subsequent regression analyses, this starting value is immaterial.

<sup>14</sup>An emerging literature considers the aggregation of TFP indices to industry-wide levels (cf. Basu and Fernald, 2002; Färe and Grosskopf, 2003; Ten Raa, 2005). Due to allocative inefficiency, the productivity of the industry is not necessarily the same as the sum of firm-level productivity levels. We are well aware of this notion, and mainly view the weighted average as *an* indicator of the industry's performance.

<sup>15</sup>The choice of activity weights has been discussed in the literature, cf. Collard-Wexler and de Loecker (2013). We believe total inputs is a better measure of activity for the demand-driven banking sector. However, all our results are robust to using total assets for computing market shares. In fact, the correlation between the two is above 0.9 in our data.

$$\begin{aligned}
(15) \quad \Delta\Pi_t = & \underbrace{\sum_{i \in S} (\Pi_{it} - \Pi_{it-1})\theta_{it-1}}_{\text{within effect}} + \underbrace{\sum_{i \in S} (\Pi_{it} - \Pi_{t-1}) (\theta_{it} - \theta_{it-1})}_{\text{between effect}} + \\
& \underbrace{\sum_{i \in N} (\Pi_{it} - \Pi_{t-1})\theta_{it}}_{\text{entry effect}} - \underbrace{\sum_{i \in X} (\Pi_{it-1} - \Pi_{t-1})\theta_{it-1}}_{\text{exit effect}}.
\end{aligned}$$

The within effect measures how firm-level changes in the survivor's productivity contribute to industry-wide performance growth. Using our earlier firm-level productivity decompositions, we can decompose the within effect further into contributions from improved technical efficiency, technical progress and increasing returns to scale. Using expression (12), we obtain:

$$(16) \quad \sum_{i \in S} (\Pi_{it} - \Pi_{it-1})\theta_{it-1} = \sum_{i \in S} (\Delta TE_{it} + TC_{it} + SC_{it})\theta_{it-1}.$$

The remaining terms in expression (15) measure the dynamics between industry participants. In a given period, firms can increase or decrease their inputs, relative to other firms, changing the share of the surviving firms. We call the resulting changes in productivity the between effect. Furthermore, entry and exit of banks can also influence productivity of the industry. Together, entry, exit and the between effect determine the size and sign of the reallocation effect. This reallocation effect will be positive if above-average productive firms grow in size (of expenditures) or enter, thereby contributing positively to industry-level productivity. Alternatively, if below-average firms decrease or exit the market, industry-wide productivity will increase as well. The reallocation effect therefore quantifies the degree of restructuring in the banking sector. Note that the within effect has a nice counterfactual interpretation: it shows what aggregate productivity would have been in the absence of reallocation. The remaining terms then measure what productivity would have been without changes in the distribution of productivity.

Summing up, we have now decomposed industry-level productivity changes into firm-level contributions and reallocation effects, stemming from industry restructuring.

### C. Data

In order to properly assess the role of reallocation in driving productivity change in the banking sectors of Europe and the US, we first need to carefully consider the sample of banks included in our study.

We only include independent banks, in order to properly measure reallocation between banks, rather than within banks or between banks that are part of the same holding company. For the US, we use the Federal Reserve Call reports

(for independent banks) and Y9-C reports (for bank holding companies), and include all banks that are either independent and not part of a bank holding company or are the largest entity of a bank holding company.<sup>16</sup> For Europe, we use Bureau van Dijk's BankScope database. Again, we select only independent banks, relying on BankScope's Independence indicator.<sup>17</sup>

The resulting data set is unbalanced as, for various reasons, not all banks are covered in the entire period. Only commercial, cooperative and savings banks are included, as these banks are more or less homogeneous in terms of financing structure. In total, we include 3,313 bank-year observations for Europe, and 6,184 for the US, where many more small banks have so far survived.

Table 1—: Descriptive statistics

	variable	Europe		US	
production set	loans ( $Y_1$ )	12,581,635	(51,754,604)	2,916,294	(24,164,706)
	investments ( $Y_2$ )	13,034,517	(77,738,824)	1,036,701	(8,774,426)
	off-balance sheet ( $Y_3$ )	177,275	(854,359)	92,532	(866,767)
	labor ( $X_1$ )	208,405	(965,436)	65,689	(565,239)
	financial capital ( $X_2$ )	699,588	(3,225,495)	104,951	(981,383)
	physical capital ( $X_3$ )	373,847	(1,711,405)	18,806	(164,015)
performance	scale elasticity	0.851	(0.014)	0.730	(0.031)
	efficiency	0.891	(0.063)	0.889	(0.056)
	productivity growth ( $\Delta TFP$ )	0.024	(0.136)	0.009	0.161
	productivity ( $\Pi$ )	2.286	0.737	2.474	0.445
	market share	0.453	(1.492)	0.243	(1.839)
	Lerner	0.457	(0.165)	0.477	(0.674)
observations		3,313		6,184	

Note:  $X_1$  is measured using personnel expenses;  $X_2$  using total interest expenses; and  $X_3$  using overhead expenses, including write-offs on physical capital. All outputs and inputs as well as operating profits in 1995 PPP dollars. Market share is in percentages.

Now that we have selected our sample, next on our agenda is the selection of each bank's production set. We follow the literature, and use the intermediation approach popularized by Sealey and Lindley (1977). We identify three outputs: loans, investments and off-balance sheet items. In addition, we identify three inputs: labor, funds and physical capital. Our main concern at this point, is to measure each item in a consistent manner across different banks in different markets. We measure all inputs and outputs in monetary terms. The inputs are measured as flows: personnel expenses, total interest expenses and overhead

<sup>16</sup>For banks that are not part of a bank holding company, we use call report items RSSD9365 and RSSD9349 to exclude those that have an outside equity holder with a majority position. For the remaining banks, item RSSD9397 is used to identify a bank as the largest entity of a bank holding company.

<sup>17</sup>We include all banks with an Independence indicator score of A or B. Banks with an indicator score of A have no shareholder with direct or total ownership exceeding 25%. Banks with a score of B can have a known recorded shareholder with an ownership above 25%, but have no shareholders with a direct or total (via cross-holdings) ownership above 50%.



expenses (including write-offs on physical capital), respectively. Two of the outputs are measured as end-of-year stocks, namely loans (net of impaired loans) and investments (consisting of all other earning assets). The remaining output, off-balance sheet items, is measured using net fees and commissions, a flow.

The full production set is described in the top part of Table 1. From the table, we observe that European banks are on average much larger. With their dense branch networks, they also spend proportionally more on physical capital. US banks, on the other hand, earn proportionally more from off-balance sheet operations.

In our analysis of the sources and consequences of productivity changes in both markets, we make use of a number of additional variables, described in the bottom part of Table 1. First, we include our estimates of banks' scale elasticity and efficiency. In terms of the notation in section III, the scale elasticity at time  $t$  is defined as  $SE = - \sum_{j=1}^J \frac{\partial \ln D_o(\mathbf{x}^t, \mathbf{y}^t, t)}{\partial \ln x_j^t}$ . Thus,  $SE = 1$  indicates constant returns, while  $SE < 1$  denote decreasing returns to scale. We observe that on average, both in Europe and in the US, banks experience decreasing returns to scale. Returns to scale are significantly lower in the US, which may play an important role in the reallocation process. Efficiency, meanwhile, is remarkably similar in both markets: an efficiency score of approximately 0.89 means that the average bank could produce 11% more outputs for a given input mix.

Reflecting the highly skewed size distribution of banks in the US, we observe that the average market share of the US banks is much lower than that of the European banks, even though concentration is higher in the former market. The market shares are calculated based on total inputs. The European banks are pooled together, so that the market share of each firm measures its size in the total European banking sector. We consider this to be the 'fair' comparison of European and US banks, as both - in principal - are supposed to operate on markets that are internally open.<sup>18</sup>

Finally, we measure banks' markup using the Lerner index, constructed as the sum of net operating income and expenses on premises and fixed assets (measuring fixed costs) over total operating income. Banks in Europe and the US appear to have, on average, very similar Lerner indices, cautiously suggesting that there are perhaps no drastic differences in the level of competition in both markets.

#### IV. Results

In this section, we describe our results. We start by describing productivity levels, and introduce our main parameter estimates and the resulting productivity measures. Next, we analyze the resulting productivity dynamics, and study

<sup>18</sup>All analyses in this paper have also been conducted for Germany, France, Italy and the United Kingdom separately. Findings as presented here are robust, and these additional results are available upon request.

how each bank market has fared over our sample period. Subsequently, we investigate the importance of reallocation in driving these dynamics. Finally, we explore the long-run dynamics in both markets.

#### *A. Productivity levels*

We start by estimating our output distance stochastic frontier model specified in equation (9) for both markets. The estimation results are summarized in Table 2. Interpreting individual coefficient estimates is notoriously difficult with these translog estimations. Instead, therefore, we describe the main features of our results by inspecting a number of key parameters.

First, we are interested in the importance of efficiency. From the estimations in Table 2, we note that  $\lambda$  is significantly larger than zero. For Europe, a  $\lambda$  of slightly more than 2 implies that the amount of residual variance that can be attributed to inefficiency is twice as large as the amount of noise. For the US, this value is somewhat lower, but still considerable, at 1.599. These inferences are confirmed by a likelihood ratio test of whether  $\sigma_u^2$ , the variance attributable to inefficiency, is significantly larger than zero, which is the case in both markets.

Second, of course, we are interested in the level of efficiency and scale economies, and their variance. After all, the potential for reallocation increases as the differences in efficiency and scale economies increase (Melitz, 2003). Figures 2a and 2b show kernel density plots of bank-level estimates of the scale elasticity in the US and Europe, respectively.

Two observations stand out: (i) the elasticity of scale is lower in the US in all sample years, and (ii) both regions witness an increase in the elasticity of scale over time. Due to both geographical and product-mix expansion, banks appear to become more able to produce outputs at given input levels.

Figures 2c and 2d show kernel density plots of bank-level estimates of efficiency in the US and Europe, respectively. For Europe, we observe a development that is similar to what we find for scale economies: over time, efficiency improves, and its distribution becomes (slightly) narrower, as efficiency differences disappear. For the US, the story is different: from the outset, the spread in efficiency is much less than is the case in Europe, and there is no clear time pattern.

In fact, Figure 2 nicely illustrates the contribution of this paper: by merely looking at aggregate numbers (or in this case, distributions), we can infer very little about the underlying dynamics that may drive industry productivity changes.

#### *B. Productivity dynamics*

We now turn to describing the evolution of productivity change in both markets. Table 3 shows the results of decomposing the year-on-year growth of

Table 2—: Frontier results

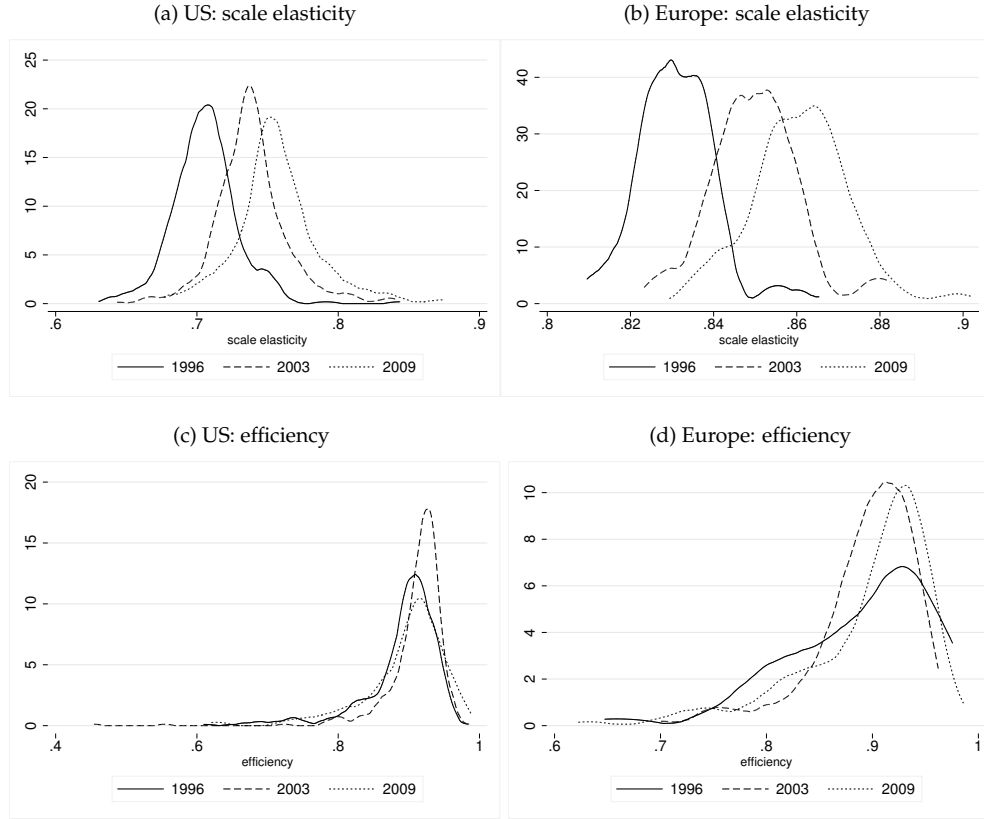
	Europe	US
$\ln y_2$	0.058 (0.032)*	0.173 (0.025)***
$\ln y_3$	0.776 (0.044)***	0.608 (0.040)***
$\frac{1}{2} \times \ln y_2 \times \ln y_2$	0.082 (0.003)***	0.039 (0.002)***
$\ln y_2 \times \ln y_3$	-0.055 (0.003)***	0.009 (0.003)***
$\frac{1}{2} \times \ln y_3 \times \ln y_3$	0.124 (0.005)***	0.050 (0.005)***
$\ln x_1$	0.333 (0.132)**	-0.154 (0.100)
$\ln x_2$	-0.194 (0.038)***	-0.200 (0.038)***
$\ln x_3$	-1.024 (0.146)***	-0.374 (0.081)***
$\frac{1}{2} \times \ln x_1 \times \ln x_1$	0.327 (0.059)***	-0.082 (0.034)**
$\ln x_1 \times \ln x_2$	0.051 (0.018)***	-0.017 (0.014)
$\ln x_1 \times \ln x_3$	-0.391 (0.047)***	0.076 (0.023)***
$\frac{1}{2} \times \ln x_2 \times \ln x_2$	-0.014 (0.006)**	0.017 (0.008)**
$\ln x_2 \times \ln x_3$	-0.035 (0.020)*	0.004 (0.011)
$\frac{1}{2} \times \ln x_3 \times \ln x_3$	0.442 (0.042)***	-0.067 (0.019)***
$\ln y_2 \times \ln x_1$	-0.092 (0.012)***	-0.001 (0.009)
$\ln y_2 \times \ln x_2$	-0.031 (0.004)***	0.013 (0.004)***
$\ln y_2 \times \ln x_3$	0.123 (0.013)***	-0.010 (0.008)
$\ln y_3 \times \ln x_1$	0.066 (0.015)***	0.019 (0.011)
$\ln y_3 \times \ln x_2$	0.033 (0.005)***	-0.006 (0.005)
$\ln y_3 \times \ln x_3$	-0.101 (0.016)***	-0.037 (0.010)***
$t$	0.012 (0.006)*	-0.026 (0.007)***
$\frac{1}{2} \times t \times t$	0.002 (0.000)***	0.004 (0.000)***
$\ln y_2 \times t$	-0.005 (0.001)***	0.003 (0.001)***
$\ln y_3 \times t$	0.009 (0.001)***	0.000 (0.001)
$\ln x_1 \times t$	-0.001 (0.004)	-0.004 (0.002)*
$\ln x_2 \times t$	0.013 (0.001)***	0.004 (0.001)***
$\ln x_3 \times t$	-0.015 (0.004)***	-0.003 (0.002)*
constant	-3.129 (0.392)***	-3.892 (0.244)***
$\lambda$	2.029***	1.599***
$LRT_{u=0}$	0.000***	0.000***
observations	3,313	6,184

Note: Dependent variable is  $-\ln y_1$ .  $\lambda = \sigma_u / \sigma_v$ , and measure the relative magnitude of inefficiency and noise.  $LRT_{u=0}$  is a likelihood ratio test, with a null hypothesis that  $\sigma_u^2 = 0$  and an alternative hypothesis that  $\sigma_u^2 > 0$ . We report the p-values corresponding to the test.

sector-wide productivity (final column labeled total) into within-firm changes (column within) and between-firm changes in market shares (column real, short for reallocation). Moreover, the reallocation effect is decomposed into its components identified in equation (15), and the within effect is decomposed into its components identified in equation (16).

The top panel of Table 3 describes the evolution of productivity in Europe. For instance, the first row shows that, between 1995 and 1996, the banking sector became 7.4% more productive, on aggregate. 4.8%, or about two-thirds, is due to banks becoming more productive, which in turn is almost fully due to technical progress (4.2%). The remaining 2.6% is due to market share reallocation. Note that, absent entry, the least productive banks gained market shares; rather,

Figure 2. : Distributions of efficiency and scale elasticities



Note: Kernel density plots, based on estimations in Table 2.

the entry of banks with a productivity premium yields a positive reallocation effect. In fact, until 2008, the between effect is typically negative, indicating that, between surviving banks, market shares are reallocated to less productive banks. Rather, it is the reshaping of the banking sector via the external margin (entry or exit) that mostly affects reallocation. We also note that the within-firm component (holding market shares fixed) is steadily contributing to productivity growth, with technical change as the most important driver.

In the US, in contrast, we observe that the internal margin plays a more important role: the between-survivors effect is mostly positive. Entrants have a productivity disadvantage during the first part of our sample, causing the reallocation effect to be mostly negative. During the first part of our sample, within-firm growth was strong, while pre-crisis, productivity levels declined. Overall, the US banking sector became more productive during the sample

Table 3—: Decomposition results

	year	within	$\Delta TE$	SC	TC	between	entry	exit	real	total
Europe	1996	0.048	0.006	0.001	0.042	-0.018	0.045	0.001	0.026	0.074
	1997	0.064	0.023	0.000	0.041	0.000	0.086	0.000	0.086	0.150
	1998	0.033	0.003	-0.004	0.034	-0.038	0.025	-0.148	0.135	0.168
	1999	0.073	0.032	0.001	0.040	-0.023	-0.019	0.000	-0.043	0.030
	2000	-0.019	-0.041	-0.012	0.034	-0.040	-0.003	0.000	-0.043	-0.062
	2001	0.006	-0.011	-0.015	0.032	-0.010	-0.038	-0.002	-0.046	-0.040
	2002	0.028	0.006	-0.011	0.032	0.019	0.054	0.029	0.044	0.072
	2003	0.057	0.020	0.004	0.034	-0.007	0.031	-0.032	0.055	0.112
	2004	0.048	0.015	-0.001	0.034	-0.043	-0.154	0.000	-0.197	-0.149
	2005	0.068	0.037	-0.010	0.041	-0.047	-0.013	0.000	-0.060	0.008
	2006	-0.037	-0.055	-0.020	0.038	0.012	-0.033	0.000	-0.021	-0.059
	2007	0.014	-0.011	-0.007	0.033	-0.001	-0.005	0.000	-0.006	0.008
US	2008	0.035	0.004	0.002	0.029	0.037	-0.003	0.004	0.030	0.065
	2009	0.040	0.002	0.008	0.029	-0.021	0.005	0.000	-0.016	0.024
	1996	0.017	-0.016	-0.016	0.049	0.001	-0.032	0.016	-0.047	-0.030
	1997	0.058	0.068	-0.069	0.058	0.024	-0.002	0.008	0.014	0.071
	1998	-0.033	-0.011	-0.069	0.047	0.004	-0.058	-0.034	-0.021	-0.054
	1999	0.021	0.003	-0.042	0.060	0.014	-0.084	-0.005	-0.065	-0.044
	2000	0.050	-0.011	0.003	0.058	0.035	-0.003	-0.001	0.033	0.083
	2001	-0.007	-0.046	-0.016	0.055	0.003	0.000	0.001	0.002	-0.005
	2002	0.063	0.011	0.000	0.052	0.008	-0.004	-0.006	0.010	0.073
	2003	0.047	-0.001	-0.002	0.050	0.018	-0.001	-0.006	0.022	0.069
	2004	0.039	0.009	-0.015	0.045	0.014	0.006	0.004	0.016	0.055
	2005	-0.073	-0.089	-0.027	0.043	0.037	0.014	-0.001	0.052	-0.021
	2006	-0.016	-0.024	-0.025	0.033	0.007	0.009	0.016	-0.001	-0.016
	2007	-0.038	-0.063	-0.005	0.031	-0.008	0.022	0.044	-0.030	-0.068
	2008	0.016	-0.002	-0.013	0.031	-0.001	0.010	-0.005	0.013	0.030
	2009	0.177	0.132	0.017	0.028	-0.001	0.000	-0.003	0.002	0.179

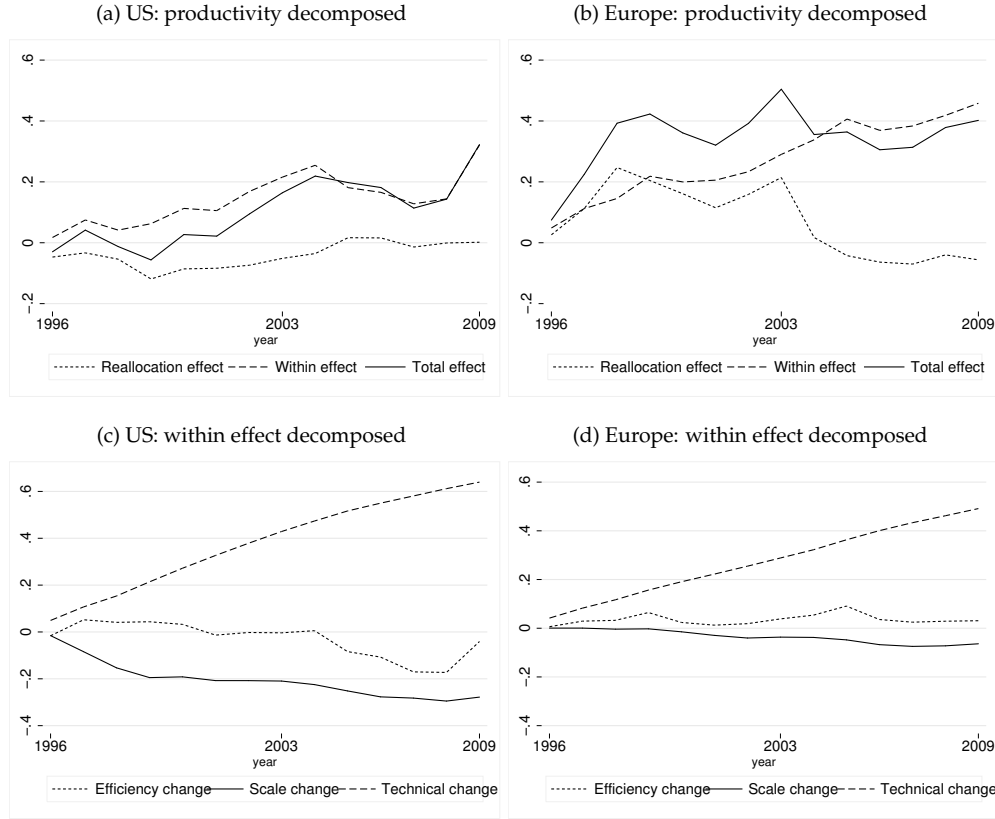
*Note:* Columns within, between, entry and exit refer to equation (15). The reallocation effect equals  $\text{real} = \text{between} + \text{entry} - \text{exit}$ . Columns  $\Delta TE$ , SC and TC refer to equation (16). The total effect equals  $\text{within} + \text{real}$ .

period.

Figures 3a and 3b plot the evolution of productivity, the within effect and the reallocation effect in the US and Europe. All graphs are cumulative effects over the sample period. Figure 3b shows more clearly how both reallocation of market shares and bank-level growth contributed to productivity growth in Europe during the first half of the sample. In later years, productivity growth slows down due to negative reallocation effects, despite continuous within-firm growth. The US banking sector is characterized by a steady rate of productivity growth, due to a strong positive productivity trend at the bank level. Market share reallocation has contributed negatively throughout, converging to zero towards the end of our sample. In total, these results imply an increase in productivity over the period 1995-2009, on the order of 11% in the US and 19% in Europe.<sup>19</sup>

<sup>19</sup>We compute these as  $(\Pi_{2009} - \Pi_{1995})/\Pi_{1995}$ .

Figure 3. : Developments in productivity



Note: Decomposition analysis based on results in Table 3.

Figures 3c and 3d document the development of the cumulative within-firm components. In both the US and Europe, technical change is the most important driver of growth, while changes in efficiency do not have any effect on growth. A main difference between the two regions concerns the returns to scale: the US banking sector is characterized by decreasing returns to scale, while slightly decreasing or constant returns to scale is more appropriate for European banks. Decreasing returns implies a less-than-proportionate increase in outputs when scaling up inputs, and hence depresses within-firm productivity growth in the US.

### C. Reallocation analysis

What are the drivers of reallocation in banking? From previous research, we know that factors such as deregulation, trade openness and product or process

innovation are important for long-term outcomes in manufacturing studies (Bartelsman and Doms, 2000). More recently, Collard-Wexler and de Loecker (2013) identify the introduction of an alternative way to produce steel, the minimill, as a major determinant of how the US steel industry looks today, by reallocating resources to the technologically superior minimill, coupled with more competition. Here, we quantify the importance for reallocation of two, related factors: bank size and market power.

The reallocation-size relationship is interesting, as it can pin down why we observe positive or negative effects of reallocation. If big banks gain market share and are more productive, reallocation will contribute to industry growth. We emphasize that no such relationship is imposed: the reallocation effect measures the contribution of *changes* in market share on productivity. We relate the bank-level reallocation effects to initial market share using a flexible fixed-effects regression. In particular, we estimate the following model separately for the US and Europe:

$$\begin{aligned}
(17) \quad & \mathbb{1}_{(i \in S_t)} [(\Pi_{it} - \Pi_{t-1}) (\Delta\theta_{it})] + \\
& \mathbb{1}_{(i \in N_t)} [(\Pi_{it} - \Pi_{t-1}) \Delta\theta_{it}] - \\
& \mathbb{1}_{(i \in X_t)} [(\Pi_{it-1} - \Pi_{t-1}) \Delta\theta_{it-1}] \\
& \equiv R_{it} = \gamma_0 + \gamma_1 \theta_{it-1} + \gamma_2 \theta_{it-1}^2 + \gamma_3 \theta_{it-1}^3 + f_i + v_{it}.
\end{aligned}$$

In equation (17),  $\mathbb{1}[\cdot]$  denotes the indicator function,  $R_{it}$  denotes a bank-level reallocation effect, obtained by summing the between effect, entry effect and exit effect,  $f_i$  denotes a bank fixed effect and  $v_{it}$  the error term. It is understood that market shares for entering and exiting banks are zero in the relevant period. We obtain this specification by testing down a flexible polynomial regression model.<sup>20</sup> Note that the fixed effect solves the initial conditions problem for converting the TFP growth series into TFP levels, as explained in section III.

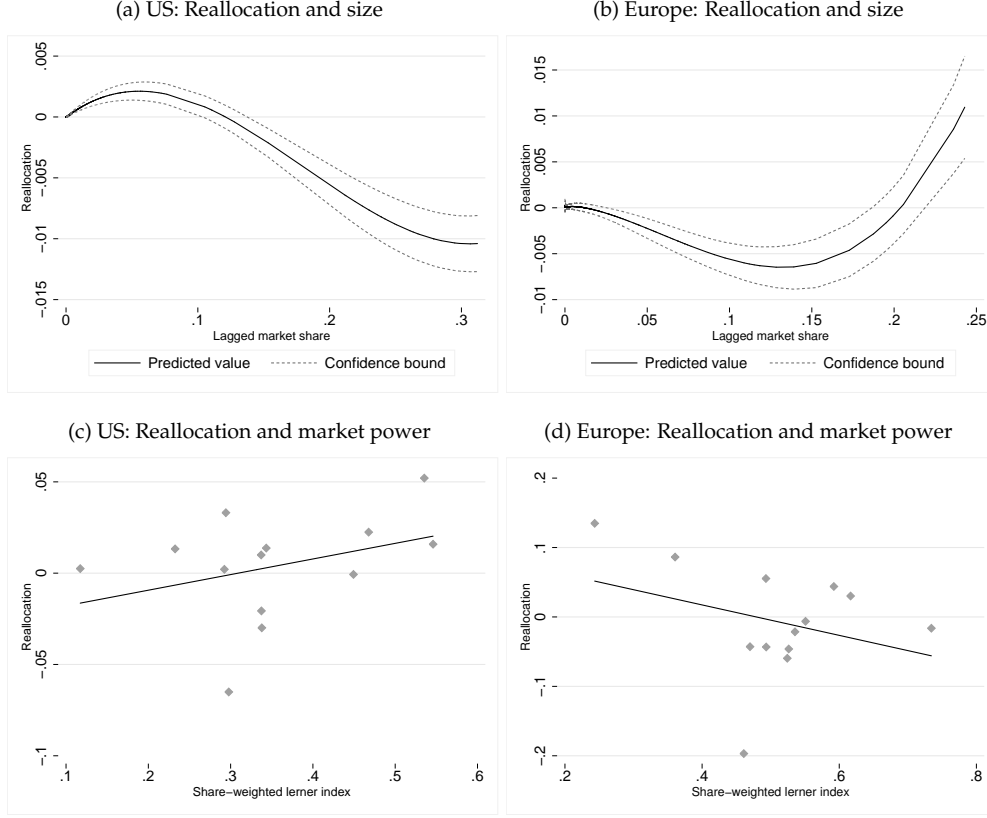
Figures 4a and 4b show the implied relationship between reallocation and lagged market share, together with a 95% confidence bound. Figure 4a clearly shows why the reallocation effect has been mostly negative in the US: large banks have lost market shares while being more productive,<sup>21</sup> driving the industry-level reallocation effect to become negative. Smaller banks have contributed positively to the reallocation effect, but the effect is smaller, yielding a negative effect.

In contrast, Figure 4b shows a very different pattern: large banks have positive reallocation effects, while small banks contribute negatively to growth. As

<sup>20</sup>Time fixed effects turned out insignificant, and are excluded from the equation.

<sup>21</sup>Inspection of the data reveals that large banks have, on average, lost, rather than gained, market share. Results available upon request.

Figure 4. : Reallocation effects



Note: Figures 4a and 4b are fitted from polynomial regression, with 95% confidence bounds. Figures 4c and 4d include linear fits based on plotted values.

Figure 3b shows, this pattern yields a positive aggregate reallocation effect during the first sample years. The difference in the reallocation-size pattern can thus explain why, on aggregate, the reshuffling of market shares contributes positively to growth in Europe during the first years, and slightly negative in the US.

Finally, we consider the reallocation-market power relationship. We use a Lerner index to measure market power. Note that the Lerner index as defined is essentially scale-free, thus not picking up the reallocation-size relationship investigated above. Collard-Wexler and de Loecker (2013) find that reallocation and competition are related, with increased competition (measured by share-weighted markups) resulting in more reallocation, at the industry level. We follow their approach, and compute a share-weighted Lerner index as  $L_t = \sum_{i=1}^N L_{it}\theta_{it}$ . Figures 4c and 4d plot the industry-level reallocation effect at time  $t$



against the share-weighted Lerner index at time  $t - 1$ , i.e. 1 observation per year. The results suggest a decreasing relationship in Europe, signalling a positive competition-reallocation nexus, while the reverse is true for the US. Obviously, the result is suggestive of an association; identifying a causal relationship is a much harder task.<sup>22</sup> Taken at face value, coupled with the finding that the share-weighted Lerner index has increased over time (paralleling the HHI, see Figure 1), the result suggests that reallocation has been lower during the second part of the sample in Europe due to an decrease in market power. Similarly, the US banking sector has gained from decreased competition over time, by stimulating reallocation of market shares towards more productive banks.

#### *D. Productivity developments over longer horizons*

Up until now, we have investigated productivity developments and its components using annual data. One issue with this approach is that short-run and long-run adjustments are mingled together: a bank restructuring its activities in an effort to become more productive may well suffer from temporary productivity losses, showing up as a negative productivity development. In other words, the gains (or losses) from restructuring which realize in later years are not attributed to reallocation. In an attempt to study productivity over longer horizons, we divide the sample into two periods, with the year 2002 as the dividing year. In particular, we pretend to observe the banking sectors only in the years 1995, 2002 and 2009. We study the growth in productivity by comparing the banking sector in 2002 with that of 1995, and similarly for the second period.<sup>23</sup> Table 4 shows the development of productivity, as well as the within and reallocation effect, over these longer horizons.<sup>24</sup>

Several observations stand out from Table 4. First, especially in the US, we observe the larger role of efficiency improvements in explaining within-firm growth. Although this may seem surprising given that the efficiency distribution appears to be stable in the US (see Figure 2), it shows the importance of looking at micro-level changes rather than focusing only on moments of distributions, an approach we have taken throughout. Technical change remains the dominant driving force in Europe. Second, in the US, the reallocation effect is positive;

<sup>22</sup>While a similar relationship could in principle be estimated using bank-level data, it requires a credible strategy for inference as the Lerner index is likely to be endogenous, as some banks may have acquired more market share by exploiting market power. Of course, identifying a causal effect of competition on reallocation is an interesting example of future research.

<sup>23</sup>We use the same estimated output distance function to calculate our measure of productivity; differences in bank-level productivity are therefore not driving these results.

<sup>24</sup>Showing cumulative reallocation effects, as we do in Figure 3, does not alleviate the problem of attributing long-run gains to within-firm growth rather than between-firm growth. Specifically, it is easy to show that the cumulative between effect for periods  $t$  and  $t + 1$  equals the long-run between effect,  $\sum_i (P_{it+1} - P_{t-1})(s_{it+1} - s_{it-1})$ , which we calculate here, minus  $\sum_i (P_{it+1} - P_{it})(s_{it} - s_{it-1})$ . The last term can be interpreted as a next-period performance effect of current changes in market share. Therefore, relative to the cumulative results, the long-run effects in Table 4 attribute these long-run gains (or losses) to the reallocation effect, rather than the within effect. We thank an anonymous referee for pointing out this deficiency in an earlier version.

Table 4—: Long run effects

	period	within	$\Delta TE$	$\Delta SC$	$\Delta TC$	real	total
EU	95-02	0.272	0.049	0.002	0.221	0.120	0.392
	02-09	0.127	-0.015	0.004	0.138	-0.118	0.010
US	95-02	0.009	-0.011	0.008	0.019	0.086	0.095
	02-09	0.156	0.118	0.015	0.039	0.072	0.228

hence, over these 7-year periods, under-performing banks lose market share. The change in the composition of the banking sectors, together with productivity improvements within surviving banks, yields a more productive sector in the US. In Europe, reallocation gives higher productivity only in the first half of our sample; this effect is undone in the second half. The results in the US are in line with the findings in manufacturing studies, see e.g. Bartelsman and Doms (2000), where the creative destruction margin is found to play a large role.

Finally, Table 4 clearly shows that most of the productivity gains in Europe are realized in the first years of our sample; in contrast, the period 2002-2009 is more influential in the US. Overall, these results show that the two banking sectors have developed rather differently over the sample period: merely observing productivity growth in both sectors hides the underlying dynamic effects which shape these industries.

## V. Conclusion

We contribute to the growing literature on the restructuring of firm assets by studying the productivity dynamics characterizing the banking markets in the EU and the US. Both markets went through a significant consolidation over the last decades, which has changed the banking sector landscape. We construct productivity indicators for banks, using an estimated stochastic output distance function to compute TFP growth. Using these indicators, we document an increase in productivity in both sectors.

Contrary to most manufacturing studies, we find a limited role for reallocation in explaining productivity growth over the sample period 1995-2009. However, reallocation does take place: under-performing banks gained market share in the US during the first sample years, while larger, more productive banks expanded in Europe during the same period. In Europe, this process reverses in later years, yielding limited effects near the end of our sample. Using a longer horizon, we document long-run benefits from reallocation in the US, but not in Europe. Within-firm productivity improvements have contributed to productivity growth throughout, with technical progress as the underlying driver.

The differences between the regions point to different market conditions under which banks operate. We identify two reasons why the reallocation pattern is different between the EU and US. First, it appears that large US

banks have lost market share to competitors, while being more productive, hence restraining productivity growth. In contrast, large European banks are both more productive and have grown, fostering growth. Second, we find a positive association between reallocation and markups in the US, and a negative association in the EU. As banks are, on average, larger near the end of our sample, reshuffling of market shares results in higher performance in the US, while lowering growth in the EU.

This paper has borrowed from a literature focusing mostly on productivity at manufacturing plants. We believe that, once appropriate adjustments are made to important concepts such as productivity, the methods can be applied to many industries, yielding interesting insights in the dynamic effects of changes in operating conditions, such as globalization, openness to trade and deregulation.

### References

- Baily, M. N., Hulten, C., and Campbell, D. (1992). Productivity dynamics in manufacturing plants. *Brookings Papers on Economic Activity: Microeconomics*, 1992:187–267.
- Bartelsman, E. J. and Doms, M. (2000). Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic Literature*, 38(3):569–594.
- Bartelsman, E. J., Haltiwanger, J. C., and Scarpetta, S. (2013). Cross-country differences in productivity: The role of allocation and selection. *American Economic Review*, 103(1):305–34.
- Basu, S. and Fernald, J. G. (2002). Aggregate productivity and aggregate technology. *European Economic Review*, 46(6):963–991.
- Berger, A. N., Demsetz, R. S., and Strahan, P. E. (1999). The consolidation of the financial services industry: Causes, consequences, and implications for the future. *Journal of Banking & Finance*, 23(2-4):135–194.
- Berger, A. N., Kashyap, A. K., Scalise, J. M., Gertler, M., and Friedman, B. M. (1995). The transformation of the U.S. banking industry: What a long, strange trip it's been. *Brookings Papers on Economic Activity*, 1995(2):55–218.
- Bertola, G. and Caballero, R. J. (1994). Irreversibility and aggregate investment. *Review of Economic Studies*, 61(2):223–246.
- Bikker, J. A. and Haaf, K. (2002). Competition, concentration and their relationship: An empirical analysis of the banking industry. *Journal of Banking & Finance*, 26(1):2191–2214.
- Bos, J. W. B., Chan, Y. L., Kolari, J. W., and Yuan, J. (2010). Competition and critical mass. METEOR Research Memorandum RM/10/063, Maastricht University.

- Bos, J. W. B. and Schmiedel, H. (2007). Is there a single frontier in a single European banking market? *Journal of Banking & Finance*, 31(7):2081–2102.
- Caballero, R. J. and Engel, E. M. R. A. (1993). Microeconomic adjustment hazards and aggregate dynamics. *Quarterly Journal of Economics*, 108(2):359–383.
- Caballero, R. J., Engel, E. M. R. A., and Haltiwanger, J. C. (1997). Aggregate employment dynamics: Building from microeconomic evidence. *American Economic Review*, 87(1):115–137.
- Caballero, R. J. and Hammour, M. L. (1996). The “fundamental transformation” in macroeconomics. *American Economic Review*, 86(2):181–186.
- Caballero, R. J. and Hammour, M. L. (1998). The macroeconomics of specificity. *Journal of Political Economy*, 106(4):724–767.
- Caves, D. W., Christensen, L. R., and Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 50(6):1393–1414.
- Coelli, T. and Perelman, S. (2000). Technical efficiency of European railways: A distance function approach. *Applied Economics*, 32(15):1967–1976.
- Collard-Wexler, A. and de Loecker, J. (2013). Reallocation and technology: Evidence from the U.S. steel industry. NBER Working Paper W18739, National Bureau of Economic Research.
- Craig, B. R. and Haubrich, J. G. (2013). Gross loan flows. *Journal of Money, Credit and Banking*, 45(2-3):401–421.
- Cuesta, R. and Orea, L. (2002). Mergers and technical efficiency in Spanish savings banks: A stochastic distance function approach. *Journal of Banking & Finance*, 26(12):2231–2247.
- Delis, M. D., Molyneux, P., and Pasiouras, F. (2011). Regulations and productivity growth in banking: Evidence from transition economies. *Journal of Money, Credit and Banking*, 43(4):735–764.
- DeYoung, R., Lang, W. W., and Nolle, D. L. (2007). How the internet affects output and performance at community banks. *The Journal of Banking & Finance*, 31(4):1033–1060.
- Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics*, 4(2):115–145.
- Dunne, T., Roberts, M. J., and Samuelson, L. (1989). Plant turnover and gross employment flows in the U.S. manufacturing sector. *Journal of Labor Economics*, 7(1):48–71.

- Ericson, R. and Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies*, 62(1):53–82.
- European Central Bank (2004). *Report on EU banking structures*. European Central Bank, Frankfurt am Main, Germany.
- European Central Bank (2006). *Report on EU banking structures*. European Central Bank, Frankfurt am Main, Germany.
- European Central Bank (2010). *Report on EU banking structures*. European Central Bank, Frankfurt am Main, Germany.
- Färe, R. and Grosskopf, S. (2003). *New Directions: Efficiency and Productivity*. Kluwer Academic Publishers, Boston, MA/London/Dordrecht.
- Färe, R., Grosskopf, S., Norris, M., and Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 84(1):66–83.
- Färe, R. and Primont, D. (1995). *Multi-output production and duality: Theory and applications*. Kluwer Academic Publishers, Boston.
- Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 120:253–281.
- Foster, L., Haltiwanger, J. C., and Krizan, C. (2001). Aggregate productivity growth: Lessons from microeconomic evidence. In *New Developments in Productivity Analysis*. NBER/Chicago University Press.
- Foster, L., Haltiwanger, J. C., and Krizan, C. (2006). Market selection, reallocation, and restructuring in the U.S. retail trade sector in the 1990s. *Review of Economics and Statistics*, 88 (4):748–758.
- Frame, W. S. and White, L. J. (2004). Empirical studies of financial innovation: Lots of talk, little action? *Journal of Economic Literature*, 42(1):116–144.
- Hopenhayn, H. (1992). Entry, exit and firm dynamics in long run equilibrium. *Econometrica*, 60(5):1127–1150.
- Jayaratne, J. and Strahan, P. E. (1996). The finance-growth nexus: Evidence from bank branch deregulation. *Quarterly Journal of Economics*, 111(3):639–670.
- Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica*, 50(3):649–670.
- Kashyap, A. K. and Stein, J. C. (2000). What do a million observations on banks say about the transmission of monetary policy? *American Economic Review*, 90(3):407–428.

- King, R. and Levine, R. (1993). Finance and growth: Schumpeter might be right. *Quarterly Journal of Economics*, 108(3):717–737.
- Klepper, S. (1996). Entry, exit, growth, and innovation over the product life cycle. *American Economic Review*, 86(3):562–583.
- Koutsomanoli-Filippaki, A., Margaritis, D., and Staikouras, C. (2009). Efficiency and productivity growth in the banking industry of central and eastern europe. *Journal of Banking & Finance*, 33(3):557–567.
- Kwast, M. L. (2007). How have the banking system and the process of financial intermediation changed? *Federal Reserve Bank of Atlanta Economic Review*, First and Second Quarters:76–82.
- La Porta, R., Lopez-De-Silanes, F., Shleifer, A., and Vishny, R. W. (1997). Legal determinants of external finance. *Journal of Finance*, 52(3):1131–1150.
- La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. W. (1998). Law and finance. *Journal of Political Economy*, 106(6):1113–1155.
- Lovell, C. (2003). The decomposition of Malmquist productivity indexes. *Journal of Productivity Analysis*, 20:437–458.
- Luttmer, E. G. J. (2007). Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics*, 122(3):1103–1144.
- Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71:1695–1725.
- Melitz, M. J. and Ottaviano, G. I. (2008). Market size, trade and productivity. *Review of Economic Studies*, 75(1):295–316.
- Mitchell, M. L. and Mulherin, J. H. (1996). The impact of industry shocks on takeover and restructuring activity. *Journal of Financial Economics*, 41(2):193–229.
- Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.
- Orea, L. (2002). Parametric decomposition of a generalized Malmquist productivity index. *Journal of Productivity Analysis*, 18:5–22.
- Park, K. and Pennacchi, G. (2009). Harming depositors and helping borrowers: The disparate impact of bank consolidation. *Review of Financial Studies*, 22(1):1–40.
- Perotti, E. C. and Suarez, J. (2002). Last bank standing: What do I gain if you fail? *European Economic Review*, 46(9):1599–1622.

- Ray, S. C. and Desli, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: Comment. *American Economic Review*, 87(5):1033–1039.
- Rose, P. S. (1987). The impact of financial services deregulation: The hypotheses and the evidence from 240 U.S. metropolitan banking markets. *Quarterly Journal of Business and Economics*, 26(2):55–88.
- Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*. Harper and Row, New York.
- Sealey, C. W., J. and Lindley, J. T. (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions. *Journal of Finance*, 32(4):1251–1266.
- Shephard, R. (1970). *Theory of Cost and Production Functions*. Princeton University Press, Princeton.
- Stiroh, K. J. and Strahan, P. E. (2003). Competitive dynamics of deregulation: Evidence from U.S. banking. *Journal of Money, Credit and Banking*, 35(5):801–828.
- Ten Raa, T. (2005). Aggregation of productivity indices: The allocative efficiency correction. *Journal of Productivity Analysis*, 24:203–209.
- Tirtiroğlu, D., Daniels, K. N., and Tirtiroğlu, E. (2005). Deregulation, intensity of competition, industry evolution, and the productivity growth of U.S. commercial banks. *Journal of Money, Credit and Banking*, 37(2):pp. 339–360.
- Wheelock, D. C. and Wilson, P. W. (1999). Technical progress, inefficiency, and productivity change in U.S. banking, 1984–1993. *Journal of Money, Credit and Banking*, 31(2):212–234.
- Wheelock, D. C. and Wilson, P. W. (2000). Why do banks disappear? The determinants of U.S. bank failures and acquisitions. *The Review of Economics and Statistics*, 82(1):127–138.
- Wheelock, D. C. and Wilson, P. W. (2012). Do large banks have lower costs? New estimates of returns to scale for U.S. banks. *Journal of Money, Credit and Banking*, 44(1):171–199.

#### APPENDIX A: THE FOSTER-HALTIWANGER-KRIZAN DECOMPOSITION

Define productivity of firm  $i$  in year  $t$  as  $\Pi_{it}$ ,  $i = 1..I$ ,  $t = 1..T$ . Let  $\theta_{it}$  denote the market share. We define industry-wide productivity as  $\Pi_t = \sum_{i=1}^{I_t} \Pi_{it}\theta_{it}$ , which equals the weighted average productivity of all firms active at time  $t$ .

The number of firms may change over time from  $I_{t-1}$  to  $I_t$ . Letting  $N$  denote entering firms,  $X$  denote exiting firms and  $S$  continuing firms, we obtain  $I_{t-1} = X + S$  and  $I_t = N + S$ . With these definitions, productivity growth between  $t - 1$  and  $t$  can be written as follows:

$$\begin{aligned}
 \Delta \Pi_t &= \Pi_t - \Pi_{t-1} = \sum_{i=1}^{I_t} \Pi_{it} \theta_{it} - \sum_{i=1}^{I_{t-1}} \Pi_{it-1} \theta_{it-1} \\
 (A.1) \quad &= \sum_{i \in S} \Pi_{it} \theta_{it} + \sum_{i \in N} \Pi_{it} \theta_{it} - \left( \sum_{i \in S} \Pi_{it-1} \theta_{it-1} + \sum_{i \in X} \Pi_{it-1} \theta_{it-1} \right)
 \end{aligned}$$

By definition, market shares must sum to 1:  $\sum_{i=1}^{I_t} \theta_{it} = 1 \forall t$ . Therefore, we also know that:

$$(A.2) \quad \left( \sum_{i \in S} \theta_{it} + \sum_{i \in N} \theta_{it} \right) - \left( \sum_{i \in S} \theta_{it-1} + \sum_{i \in X} \theta_{it-1} \right) = 0.$$

From equation A.1, we add and subtract:

$$(A.3) \quad \sum_{i \in S} \Pi_{it} \theta_{it-1},$$

and we subtract:

$$(A.4) \quad \Pi_{t-1} \left( \left( \sum_{i \in S} \theta_{it} + \sum_{i \in N} \theta_{it} \right) - \left( \sum_{i \in S} \theta_{it-1} + \sum_{i \in X} \theta_{it-1} \right) \right),$$

to obtain the following:

$$\begin{aligned}
 \Delta \Pi_t &= \sum_{i \in S} \Pi_{it} \theta_{it} + \sum_{i \in N} \Pi_{it} \theta_{it} - \sum_{i \in S} \Pi_{it-1} \theta_{it-1} - \sum_{i \in X} \Pi_{it-1} \theta_{it-1} + \sum_{i \in S} \Pi_{it} \theta_{it-1} \\
 &\quad - \sum_{i \in S} \Pi_{it} \theta_{it-1} - \Pi_{t-1} \left( \left( \sum_{i \in S} \theta_{it} + \sum_{i \in N} \theta_{it} \right) - \left( \sum_{i \in S} \theta_{it-1} + \sum_{i \in X} \theta_{it-1} \right) \right) \\
 (A.5) \quad &= \sum_{i \in S} (\Pi_{it} - \Pi_{it-1}) \theta_{it-1} + \sum_{i \in S} \Pi_{it} (\theta_{it} - \theta_{it-1}) - \Pi_{t-1} \sum_{i \in S} (\theta_{it} - \theta_{it-1}) \\
 &\quad + \sum_{i \in N} (\Pi_{it} - \Pi_{t-1}) \theta_{it} - \sum_{i \in X} (\Pi_{it-1} - \Pi_{t-1}) \theta_{it-1} \\
 &= \sum_{i \in S} (\Pi_{it} - \Pi_{it-1}) \theta_{it-1} + \sum_{i \in S} (\Pi_{it} - \Pi_{t-1}) (\theta_{it} - \theta_{it-1}) \\
 &\quad + \sum_{i \in N} (\Pi_{it} - \Pi_{t-1}) \theta_{it} - \sum_{i \in X} (\Pi_{it-1} - \Pi_{t-1}) \theta_{it-1},
 \end{aligned}$$



which is equal to equation (15). For this decomposition, productivity growth is due to within-firm productivity growth evaluated at initial market shares (the first term), market share reallocation towards above-average productive firms (the second term), entrants with productivity above the previous period's average productivity (the third term) and due to below-average firms exiting the market (the fourth term). This decomposition is due to Foster, Haltiwanger and Krizan (2001).